Hybrid projection estimation for a wide class of functional parameters

Jean-Baptiste Aubin, Rosaria Ignaccolo

Hybrid projection estimation for a wide class of functional parameters

Jean-Baptiste Aubin[*]

INSA-Lyon, ICJ,
20, Rue Albert Einstein, 69621 Villeurbanne Cedex, France
e-mail: jean-baptiste.aubin@insa-lyon.fr

and

Rosaria Ignaccolo

Dipartimento di Economia e Statistica "Cognetti de Martiis",
Campus Luigi Einaudi, Università di Torino, Italy
e-mail: rosaria.ignaccolo@unito.it

Abstract:
In the framework of functional parameters estimation (such as e.g. density estimation), we consider a wide class characterized by the fact that its elements can be written as limits of sums of the expected values of random variables. We propose an "hybrid" projection estimator of such a general functional parameter when we observe n realizations of a discrete time stochastic process (X_t). The estimator is said "hybrid" because the dimension of the projection subspace is chosen differently according to the sample size, very large or not. From the asymptotic point of view, this estimator locally reaches a superoptimal rate for the mean integrated square error (MISE) on a dense subset of the space to which the considered functional parameter is supposed to belong, and we state under which hypotheses there is a near-optimal rate of convergence elsewhere in L^2. Note that some hypothesis have been relaxed with respect to previous literature. The finite sample performance is clearly improved with respect to other estimators of the same kind; indeed performance is evaluated through a simulation study, where the parameter to estimate is the spectral density: the proposed estimator is shown to reduce often drastically the MISE in comparison with that of the classical projection estimator and the kernel estimator. Finally, from the practioner point of view this new estimator can be completely data-driven with only a smoothing parameter to choose.

1. Introduction

Many papers and books in statistical literature treat the estimation of functional parameters considering one at a time for example the density, the regression function or the spectral density, whereas a more general approach is possible. We adopt here the projection estimation introduced by Cencov (1962), that consists in projecting a potentially infinite dimensional functional parameter φ onto a subspace with finite dimension k_n (that increases with the size n of the observed sample) and then estimating its projection. By assumption the

[*]Corresponding author

functional parameter admits a Fourier development through a given projection basis, so that a natural way for its estimation consists in truncating the series and estimating a finite number of Fourier coefficients. After Cencov's paper, several results enrich the theory about the projection estimation. Bleuez and Bosq (1976) give necessary and sufficient conditions for the density projection estimator to be consistent. Delecroix and Protopopescu (2001) define limits for the Mean Integrated Square Error (MISE) for regression; these limits are particularized in the cases of the trigonometric basis, Legendre basis and Haar basis, so that rates of convergence of the MISE are given. Devroye and Gyorfi (1985) and Devroye (1987) evaluate density estimators by a L_1 norm criterion and Efron and Tibshirani (1996) try to solve problems linked to the fact that the density projection estimator is not necessarily positive.

The projection estimator is based on the choices of a projection basis and of a truncation index. Of course, both of them play a central role in the behaviour of the estimator. As a matter of fact, several authors cope with these issues. To choose a basis, Efromovich (1999) in the case of the density and Eubank and Speckman (1990) in the case of the regression both use results of Krylov (1955) concerning orthonormal systems. Other basis examples can be found in Sansone (1959), Kolmogorov and Fomin (1957) or Antoniadis (2007). Instead, the truncation index (or cutoff), that is the projection subspace dimension, is usually chosen to minimize an estimation of the MISE (see Hart 1985; Diggle and Hall 1986; or Tarter and Lock 1993). Nevertheless, the quantity to minimize is the sum of an infinite series and this minimization problem can be solved in an iterative way, considering the contribution of each term added to estimate the MISE (Kronmal and Tarter 1968). This procedure has been generalised by Bosq and Lecoutre (1987) and by Aubin and Leoni-Aubin (2008) in a semiparametric framework.

In the last decades, new estimators are presented such as those based on kernels or wavelets (see for example Prakasa-Rao 1983; Silverman 1986; Nadaraya 1989; Thompson and Tapia 1990; Hart 1997; Efromovich 1999; or more recently Bosq and Blanke 2007). In the wavelets' framework, Donoho *et al.* (1996) and Lepski, Mammen and Spokoiny (1997) introduce two ways of choosing which coefficients are kept in the estimation procedure, using respectively "hard" and "soft" thresholding. The soft thresholding consists in keeping all the coefficients associated to the first components of the basis until a certain k_n (chosen by the statistician). The hard threshold method only keeps the Fourier terms associated with estimated coefficients larger than a certain threshold, and discards terms with small estimated coefficients despite their order in the Fourier development. According to Donoho *et al.* (1996), "*one expects that soft thresholding will better suppress noise artifacts, while hard thresholding will better preserve the visual appearance of peaks and jumps*". Picard and Tribouley (2000) demonstrate that soft thresholding estimators give better confidence intervals than those constructed through hard thresholding. In the orthogonal series' framework, Efromovich (1996, 1999, 2010) introduces a "universal" estimator which combines hard and soft thresholding procedures. Instead, Bosq (2002,2005) introduces an "adaptive" version of the projection estimator in the case of density.

The adaptivity lies in the choice of the dimension of the projection space. The chosen dimension $\widehat{k}_n$ corresponds in most cases to the greatest $j \leq k_n$ such that the absolute value of the estimated Fourier coefficient is greater than a threshold

$$\gamma_n := c\sqrt{\frac{\log(n)}{n}} \tag{1}$$

where c and k_n are fixed by the statistician and n denotes the sample size. A strong advantage of this method is the fact that $\widehat{k}_n$ is a consistent estimator of the real order of development of ϕ with respect to the projection basis. If this real order is finite then the "adaptive" estimator is superoptimal with respect to the MISE. Some assumptions (as the independence of data) have been relaxed in Aubin (2006) in the case of the density, and in Souare (2008) in the case of the spectral density.

Subsequently, Bosq and Blanke (2007) consider a wide class of functional parameters introduced by Carbon (1984), characterized by the fact that the elements of this class can be written as limits of sums of the expected values of random variables. The Carbon's class contains density, numerator of regression, spectral density, among others. Theoretical properties of the ("adaptive" and not) projection estimators for this class are presented in Bosq and Blanke (2007, Section 3.7). However, their approach is not totally satisfactory: the threshold for the estimated Fourier coefficients is completely specified as a function of n and it also depends on a multiplicative constant c (see (1)) to choose in a suitable way in order to fulfill assumptions related to asymptotic properties. Having finite samples, so far the user had to choose this "suitable constant" in the threshold that appears in the definition of the data driven truncation index $\widehat{k}_n$ (that estimates the projection subspace dimension). Aubin (2005) shows the tremendous influence of this constant on the behaviour of $\widehat{k}_n$ and consequently on the projection estimator itself through simulations.

The main drawback of all the thresholding methods is that c has to be well chosen - with respect to the data - to allow a sharp estimation. For example, Efromovich (1999, 2010) recommend to use $c = \sqrt{2}$ as a standard choice, while Bosq and Blanke (2007) consider a sequence only depending on n growing slowly to infinity. Although these choices make the procedure more easy-to-use, they clearly do not take into account a large part of available information (e.g. by means of estimated coefficients) and cannot provide a sharp-enough choice of c to give completely convincing results for a finite sample.

In this work, we propose an *"hybrid" projection estimator (hPE)* both estimating very accurately φ even for small sample sizes and satisfying the asymptotic superoptimality with respect to the MISE when φ admits a finite development. These properties hold for any value of c, so relaxing stronger constraints in previous literature. The hPE is called "hybrid" because it is featured by the fact that the threshold is defined as the minimum between two quantities: (i) a data-driven component, that does not require any choice of tuning parameters but the maximum dimension of the projection subspace k_n; (ii) a quantity dependent on

n up to some constants (this latter quantity coincides with the threshold defined in Bosq and Blanke, 2007, in the case of geometrically $\alpha - mixing$ processes). We will consider dependent data generated by $\alpha - mixing$ processes in two cases: the associated $\alpha - mixing$ coefficients are geometrically or arithmetically decreasing.

The rest of the paper is as follows. In Section 2 we recall the definition of the Carbon's class of functional parameters and the hybrid projection estimator, and we set the assumptions for the theoretical framework. In Section 3, we analyse the large sample behaviour of the proposed estimator hPE and its associated truncation index $\widehat{k}_n$ that is, under mild hypotheses, an estimator of the right order of the development of φ with respect to the projection basis. Also, we show that hPE reaches a superoptimal rate for MISE on a dense subset $\mathcal{G}_0$ of the space H to which φ belongs and we state under which conditions this estimator reaches a near-optimal rate of convergence when the unknown parameter φ does not belong to $\mathcal{G}_0$. Note that these results do not depend on c. In Section 4 we illustrate the estimator finite sample behaviour in the case of the spectral density estimation. Simulation results allow to evaluate the finite sample performance of $\widehat{k}_n$ and hPE (using the MISE criterion), and to compare hPE to the classical projection estimator and to the kernel estimator. Concluding remarks are given in Section 5, whereas all the proofs are postponed in the Appendix.

2. Hybrid Projection Estimator (hPE)

Let consider n observations $\{x_1, \dots, x_n\}$ from a discrete time stochastic process $\{X_t\}_{t\in\mathbb{Z}}$ where the random variables X_t are defined on $(\Omega, \mathcal{A}, \mathbb{P})$ and take values in a measurable space $(E, \mathcal{B})$. We are interested in the estimation of a general parameter φ, that depends on the distribution $\mathbb{P}_X$ of the process $\{X_t\}_{t\in\mathbb{Z}}$, that is $\varphi = g(\mathbb{P}_X)$, and could have an infinite dimension, belonging to a separable real Hilbert space H, equipped with scalar product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$.

The class of functional parameters introduced by Carbon (1984) can be formalized in the following definition (as in Bosq and Blanke, 2007).

We will say that φ is ***(e,h)–adapted*** if there exists a fixed orthonormal system $e = (e_j,\, j \geq 0)$ of H and a family $h = (h_j,\, j \geq 0)$ of applications $h_j : E^{\nu(j)+1} \longrightarrow \mathbb{R},\ j \geq 0$ such that

$$\varphi = \sum_{j=0}^{\infty} \varphi_j e_j \tag{2}$$

with

$$\varphi_j = \langle \varphi, e_j \rangle = \mathbb{E}\left[h_j\left(X_0, \dots, X_{\nu(j)}\right)\right], \quad j \geq 0,\ \varphi \in H,$$

where $\mathbb{E}\left[h^2\left(X_0, \dots, X_{\nu(j)}\right)\right] < \infty$ and $\nu(j) \leq j, \quad j \geq 0$.

Density, numerator of regression, spectral density, and other functional parameters are (e, h)–adapted, under suitable conditions, as shown in Bosq and Blanke (2007). For them, the functions h_j are specified by writing down the Fourier development (2) whereas the orthonormal system e can be chosen by the statistician, even if sometimes a natural choice is given, as in the case of the spectral density (where e is the cosine basis).

The estimation procedure proposed here holds for the class

$$\mathcal{C} := \{\varphi \in H : \varphi \text{ is } (e,h)\text{–adapted}\}.$$

Some technical hypotheses hold in the following.
First, the functions h_j need to satisfy the following condition for every j:

$$\exists M < \infty \qquad \text{such that} \qquad \|h_j\|_\infty < M;$$

note that this hypothesis can be satisfied by convenient conditions either on the random variables X_t or on the functions e_j of the projection basis (e.g. the trigonometric basis verifies this condition).
We also suppose that the random variables

$$(Y_j)_i := h_j(X_i, \dots, X_{i+\nu(j)}) - \mathbb{E}\left(h_j(X_i, \dots, X_{i+\nu(j)})\right)$$

are α–mixing (w.r.t. i), according to the Rosenblatt (1956) definition, and in the following we deal with two cases:

(A) arithmetically mixing: $\exists \delta > 0$ such that $\forall n \in \mathbb{N}, \alpha(n) \le n^{-\delta}$,
(G) geometrically mixing: $\exists a > 0, b > 0$ such that $\forall n \in \mathbb{N}, \alpha(n) \le a \exp(-bn)$,

where $\alpha(n)$ are the mixing coefficients.

The ***hybrid projection estimator*** **(hPE)** for $\varphi \in \mathcal{C}$ is defined by

$$\widehat{\varphi}_n := \sum_{j=0}^{\widehat{k}_n} \widehat{\varphi}_{j,n} e_j \tag{3}$$

where $\widehat{\varphi}_{j,n} = \frac{1}{n-\nu(j)} \sum_{i=1}^{n-\nu(j)} h_j(X_i, \dots, X_{i+\nu(j)})$ is an empirical unbiased estimator of φ_j and the truncation index $\widehat{k}_n$ is defined by

$$\widehat{k}_n = \begin{cases} \max\{0 \le j \le k_n : |\widehat{\varphi}_{j,n}| > \gamma_n\} & \text{if } \{0 \le j \le k_n : |\widehat{\varphi}_{j,n}| > \gamma_n\} \neq \emptyset \\ k_n & \text{if } \{0 \le j \le k_n : |\widehat{\varphi}_{j,n}| > \gamma_n\} = \emptyset \end{cases} \tag{4}$$

where the statistician chooses the sequence $(k_n)_{n\in\mathbb{N}}$ such that $k_n < n$, $k_n \to \infty$, $k_n/n \to 0$, while the threshold is defined as follows:

$$\gamma_n = \min\left(\sqrt{\frac{\sum_{j=0}^{k_n} \widehat{\varphi_{j,n}}^2}{k_n+1}}, c\sqrt{\frac{\log(n)}{n^\beta}}\right) \text{ with } 0 < \beta < 1 \text{ for the case } \textbf{(A)},$$

$$\gamma_n = \min\left(\sqrt{\frac{\sum_{j=0}^{k_n} \widehat{\varphi_{j,n}}^2}{k_n+1}}, c\sqrt{\frac{\log^{2\Gamma}(n)}{n}}\right) \text{ with } \Gamma > 1 \text{ for the case } \textbf{(G)}, \tag{5}$$

with $c \geq 0$.

If $\{0 \leq j \leq k_n : |\widehat{\varphi}_{j,n}| > \gamma_n\} = \emptyset$, it means that all the estimates of the Fourier coefficients have absolute value smaller than γ_n. Hence, according to the definition of $\widehat{k}_n$, we come back to the "classical" (*à la* Cencov) projection estimator. This happens also when $c = 0$ since $\gamma_n = c = 0$ and hence $\widehat{k}_n = k_n$.

The definition of γ_n is fundamental because a good threshold allows for a good behaviour of $\widehat{k}_n$ and of the resulting estimator $\widehat{\varphi}_n$ (see the following Section). The thresholding is operated through γ_n and the research for a maximum in (4) stops at j^* when it does not exist an integer $j_0 > j^*$ such that the absolute value of $\widehat{\varphi}_{j_0,n}$ is larger than the threshold γ_n. Thus we keep all the $\widehat{\varphi}_{j,n}$ with $j \leq \widehat{k}_n$ (so that we also keep some coefficients smaller than the threshold), while "hard" thresholding keeps only the estimated Fourier coefficients larger - in absolute value - than the threshold.
The threshold introduced here is "hybrid" in the sense that it can assume a different nature according to the sample size n, the dimension k_n and other constants. If $\gamma_n = c\sqrt{\frac{\log^{2\Gamma}(n)}{n}}$ in the case **(G)**, we go back to the projection estimator in Bosq and Blanke (2007). We have an analogous estimator in the case **(A)** if $\gamma_n = c\sqrt{\frac{\log(n)}{n^\beta}}$. Nevertheless, in such cases the threshold γ_n depends not only on c but also on Γ in the case **(G)** and β in the case **(A)** that are unknown (see Section 4 for a brief discussion).
However, if $\gamma_n = \sqrt{\frac{\sum_{j=0}^{k_n} \widehat{\varphi_{j,n}}^2}{k_n+1}}$ we do not need to know any constant. In this case, the truncation index $\widehat{k}_n$ reaches the biggest integer j such that $\widehat{\varphi}_{j,n}^2$ is larger than the average of the first $(k_n + 1)$ squared estimated Fourier coefficients. Thus the research of a maximum in the definition (4) stops on the rank of the last squared estimated coefficient larger than this average. Consequently the truncation index is such that the discarded estimated coefficients are expected to be small, and therefore the hPE estimator should keep the most part of the available information.

However, for any value of c if n is large enough, we will see in Proposition 3.1. that, under mild conditions, almost surely $\gamma_n = c\sqrt{\frac{\log^{2\Gamma}(n)}{n}}$ in the case **(G)** and $\gamma_n = c\sqrt{\frac{\log(n)}{n^\beta}}$ in the case **(A)**, respectively. Thus the hybrid projection estimator for large sample maintains good properties as in Bosq and Blanke (2007).
Instead, for finite samples the application of the hybrid projection estimator does not ask for precising any constant (except k_n) and does not need to establish if data fits in the case **(G)** or **(A)**, as we will see in Section 4.

3. Asymptotic behaviour of hPE

The asymptotic behaviour of the considered estimator differs in the case when φ admits a finite development or not, with respect to the basis e, that is when $\varphi \in \mathcal{G}_0$ or $\varphi \in \mathcal{G}_1$ where

$$\begin{aligned}
\mathcal{G}_0(0) &:= \{\varphi \in H : \varphi_j = 0, \quad \forall j > 0\}, \\
\forall K \geq 1, \mathcal{G}_0(K) &:= \{\varphi \in H : \varphi_K \neq 0 \text{ and } \varphi_j = 0, \quad \forall j > K\}, \\
\mathcal{G}_0 &:= \bigcup_{K \in \mathbb{N}} \mathcal{G}_0(K), \\
\mathcal{G}_1 &:= H - \mathcal{G}_0.
\end{aligned}$$

We show here that the hybrid projection estimator defined in (3) reaches for the MISE a superoptimal rate on $\mathcal{G}_0$. Cencov (1962) showed that the rate of the classical projection estimator is of order $\frac{k_n}{n}$ where $k_n \to \infty$ with n. The thresholding projection estimator achieves a rate of order $\frac{1}{n}$. Instead, over $\mathcal{G}_1$ we have a little loss in the convergence rate, precisely of order $\log^{2\Gamma} n$ $(\Gamma > 1)$ for the case **(G)** and of order $n^{1-\beta} \log n (0 < \beta < 1)$ for the case **(A)**; for this reason we say that hPE reaches a "near-optimal" rate in case **(G)**. These results are demonstrated for n large enough thanks to the fact that almost surely the hybrid threshold γ_n in (5) takes values according to the not data-driven case, as shown in the following proposition.

Proposition 3.1. *In the case* **(A)**, *if* $\exists j_0 : \varphi_{j_0} \neq 0$, $k_n = o\left(\frac{n^\beta}{\log(n)}\right)$ *and* $\delta > \frac{1+\beta}{1-\beta}$ *then* $\forall c > 0$ *it exists* n_c *such that for* $n > n_c$ *we have*

$$\sqrt{\frac{\sum_{j=0}^{k_n} \widehat{\varphi_{j,n}}^2}{k_n + 1}} > c\sqrt{\frac{\log(n)}{n^\beta}} \qquad \textit{almost surely.}$$

In the case **(G)**, *if* $\exists j_0 : \varphi_{j_0} \neq 0$, $k_n = o\left(\frac{n}{\log^{2\Gamma}(n)}\right)$, *then* $\forall c > 0$ *it exists* n_c *such that for* $n > n_c$ *we have*

$$\sqrt{\frac{\sum_{j=0}^{k_n} \widehat{\varphi_{j,n}}^2}{k_n + 1}} > c\sqrt{\frac{\log^{2\Gamma}(n)}{n}} \qquad \textit{almost surely.}$$

3.1. Superoptimality in $\mathcal{G}_0$

In order to analyse the asymptotic behaviour of the proposed hPE estimator when $\varphi \in \mathcal{G}_0$, first we study $\widehat{k}_n$ defined in (4): the following property shows that $\widehat{k}_n$ behaves like an estimator of K, that is the order of development of φ with respect to the basis.

Proposition 3.2. *If it exists* K *such that* $\varphi \in \mathcal{G}_0(K)$ *and*

- *in the case* **(A)** *if* $k_n = o\left(\frac{n^\beta}{\log(n)}\right)$ *and if* $\delta > \frac{2+5\beta/4}{1-\beta}$,

- *in the case* **(G)** *if* $k_n = o\left(\frac{n}{\log^{2\Gamma}(n)}\right)$,

then for n *large enough*

$$\widehat{k}_n = K \ a.s.$$

The finite development of φ and the previous proposition allow to prove the following theorem.

Theorem 3.3. *If it exists* K *such that* $\varphi \in \mathcal{G}_0(K)$ *and*

- *in the case* **(A)** *if* $k_n = o\left(\frac{n^\beta}{\log(n)}\right)$ *and if* $\delta > \frac{3+5\beta/4}{1-\beta}$,
- *in the case* **(G)** *if* $k_n = o\left(\frac{n}{\log^{2\Gamma}(n)}\right)$,

then $n\mathbb{E}\|\widehat{\varphi}_n - \varphi\|^2 = \mathcal{O}(1)$

By recalling that $MISE(\widehat{\varphi}_n) = \mathbb{E}\|\widehat{\varphi}_n - \varphi\|^2$, we have $nMISE(\widehat{\varphi}_n) = \mathcal{O}(1)$. Thus the estimator hPE reaches a rate of convergence for MISE of order $\frac{1}{n}$ that is a parametric rate.

3.2. Convergence in $\mathcal{G}_1$

If φ belongs to $\mathcal{G}_1$, the data driven truncation index $\widehat{k}_n$ tends to infinity with n, as stated in the following property.

Proposition 3.4. *If* $\varphi \in \mathcal{G}_1$ *and*

- *in the case* **(A)** *if* $k_n = o\left(\frac{n^\beta}{\log(n)}\right)$ *and if* $\delta > \frac{2+5\beta/4}{1-\beta}$,
- *in the case* **(G)** *if* $k_n = o\left(\frac{n}{\log^{2\Gamma}(n)}\right)$,

then

$$\widehat{k}_n \to \infty \ a.s.$$

In $\mathcal{G}_1$, it is also possible to derive the convergence rate of $MISE(\widehat{\varphi}_n)$ in the following theorem.

Theorem 3.5. *If* $\varphi \in \mathcal{G}_1$ *and*

- *in the case* **(A)** *if* $k_n = o\left(\frac{n^\beta}{\log(n)}\right)$ *and if* $\delta > \frac{2+5\beta/4}{1-\beta}$, *then*

$$\mathbb{E}\|\widehat{\varphi}_n - \varphi\|^2 = \mathcal{O}\left(\frac{k_n \log n}{n^\beta}\right);$$

- *in the case* **(G)** *if* $k_n = o\left(\frac{n}{\log^{2\Gamma}(n)}\right)$, *then*

$$\mathbb{E}\|\widehat{\varphi}_n - \varphi\|^2 = \mathcal{O}\left(\frac{k_n \log^{2\Gamma} n}{n}\right).$$

We recall that in the case **(G)** Bosq and Blanke (2007) proved the same result about a quasi-optimal rate with c depending on n and growing slowly to infinity, whereas we relax this assumption since the previous theorem holds for all c.

4. Finite sample behaviour of hPE

In order to analyse the finite sample behaviour of the hPE estimator, we look at the behaviour of the truncation index and we evaluate $MISE(\widehat{\varphi}_n)$ in a simulation study. Moreover we compare hPE with the classical projection estimator and the kernel one by examining their MISEs.

We focus on a particular (e,h)-*adapted* functional parameter, that is the spectral density of a discrete time zero-mean stationary stochastic process. In this case, the choice of the cosine system as projection basis is naturally imposed by writing

$$f(\lambda) = \frac{1}{2\pi}\sum_{t\in\mathbb{Z}} c_t cos\lambda t = \sum_{j\in\mathbb{N}} p_j c_j e_j(\lambda) \qquad \text{for} \qquad \lambda \in [-\pi,\pi]$$

where

$$p_j = \begin{cases} \dfrac{1}{\sqrt{2\pi}} & \text{if} \quad j=0 \\ \dfrac{1}{\sqrt{\pi}} & \text{if} \quad j\neq 0 \end{cases}$$

and c_j represents the autocovariances $c_j = \mathbb{E}\left(X_0 X_j\right),\ j\geq 0$, such that $\sum_{j=0}^{\infty}|c_j| < \infty$, and the basis is given by $e_0(\lambda) = \frac{1}{\sqrt{2\pi}}$ and $e_j(\lambda) = \frac{cos\lambda j}{\sqrt{\pi}}$.
So the functional parameter is $\varphi = f(\lambda) = \sum_{j\in\mathbb{N}} p_j c_j e_j(\lambda)$ with $\varphi_j = p_j c_j$ and the estimator of φ_j is given by

$$\widehat{\varphi}_{j,n} = \frac{1}{n-j}\sum_{i=1}^{n-j} h_j(X_i,\ldots,X_{i+j}) = \frac{1}{n-j}\sum_{i=1}^{n-j} p_j X_i X_{i+j}.$$

Since we have two kinds of asymptotic behaviour for the MISE, a superoptimal rate when φ admits a finite development and a not optimal rate when it belongs to $\mathcal{G}_1$, we consider two processes to generate observations in order to reproduce a favourable case for hPE and a not favourable one.
Specifically, if we consider n observations from a $MA(q)$ process, the spectral density belongs to $\mathcal{G}_0(q)$ since it admits a finite development of order q. Hence the estimator $\widehat{k}_n$ is an estimator for q, the true order of the Moving Average process, and from the theoretical results in Section 3 we know that it is strongly consistent. Actually we consider a $MA(1)$ process

$$X_t = Z_t + \theta Z_{t-1}$$

where Z_t is a white noise with $\sigma^2 = Var(Z_0)$, so that its spectral density is

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|1+\theta\exp^{-i\lambda}\right|^2 = \frac{\sigma^2}{2\pi}(1+2\theta\cos(\lambda)+\theta^2).$$

On the other hand, if the spectral density does not admit a finite development we can consider $MA(\infty)$ process, that is an $AR(1)$ process:

$$X_t = \phi X_{t-1} + Z_t$$

where Z_t is a white noise with $\sigma^2 = Var(Z_0)$, and its spectral density is

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|1-\phi\exp^{-i\lambda}\right|^{-2} = \frac{\sigma^2}{2\pi}(1-2\phi\cos(\lambda)+\phi^2)^{-1}.$$

In our simulation study, we consider $\theta = 0.05, 0.20, 0.35, 0.50, 0.65, 0.80, 0.95$ in the MA case and the same values of ϕ in the AR case. Moreover, we consider the innovations variance σ^2 taking the set of values $\{0.16, 0.36, 0.64, 1, 1.44, 1.96\}$. For sample sizes $n = 50, 80, 100, 150$, we generate 1000 independent sets of samples. For the sake of brevity, the following Figures show results for $n = 50$ and $n = 150$.

First of all, our simulations are devoted to explore the finite sample behaviour of $\widehat{k}_n$ and of $MISE(\widehat{\varphi}_n)$ both for $MA(1)$ and $AR(1)$ processes. The definition of $\widehat{k}_n$ strongly depends on γ_n (see (4)) and indeed we choose a constant c large enough to have $\gamma_n = \sqrt{\frac{\sum_{j=0}^{k_n}\widehat{\varphi_{j,n}}^2}{k_n+1}}$. Indeed, to compare the two quantities in the definition (5) of γ_n we would need to know first of all if data are drawn from a geometrically (or arithmetically) α-mixing process, and then the associated decreasing rate constants δ (or a and b) in order to fix values for $\Gamma > 1$ (or $0 < \beta < 1$). Unfortunately, so far the literature about the estimation of mixing coefficients is very limited; to the best of our knowledge, only McDonald *et al.* (2015) propose an estimator for the β-mixing coefficients based on a single stationary sample path but not for the α-mixing case. Hence we cannot take information from data to fix a value for Γ (or β). As explained before, the choice of a very large c has the advantage to make γ_n data driven so that the user does not need to fix other constants but k_n.

In practice, the three estimators we want to compare ask for a choice: a "smoothing parameter". With the proposed data driven threshold we need to choose only k_n for hPE, as in the case of the truncation index for the classical projection estimator (PE); for both of them we take k_n as given in (6). Instead for the kernel estimator, the selection of the smoothing parameter corresponds to the choice of the bandwidth.

Concerning the choice of the maximum dimension of the projection space, $k_n + 1$, in order to satisfy the conditions $k_n < n$, $k_n \to \infty$, $k_n/n \to 0$, and the assumptions in Proposition 3.1 (for both the cases (A) and (G)) we choose the following sequence k_n:

$$k_n := 2\lfloor\log(n)\rfloor + 1 \tag{6}$$

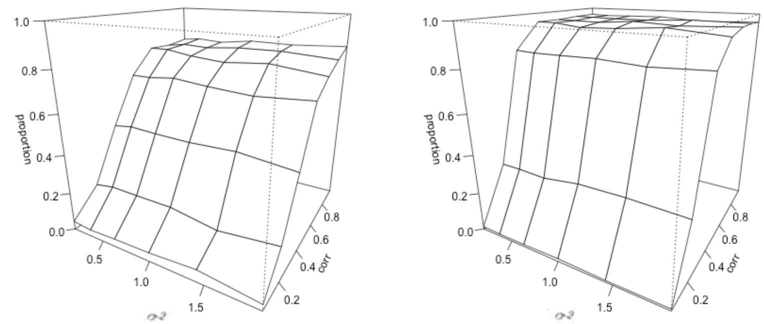

FIG 1. *Proportion of the occurrences* $\{\widehat{k}_n = 1\}$ *in the* $MA(1)$ *case, with* $n = 50$ *(left) and* $n = 150$ *(right)*

where $\lfloor x \rfloor$ is the integer part of x. As explained in Aubin (2005), a sequence of (k_n) increasing slowly to infinity allows to reduce both the number of estimated parameters in the model and the computation time.

4.1. $MA(1)$ *process*

When we have observations generated by a $MA(1)$ process, the spectral density admits a finite development with $K = 1$ and hence we are in a very favourable case for the proposed estimator hPE. Since the truncation index $\widehat{k}_n$ estimates the parameter K, we evaluate the proportion (that is the relative frequency) of the occurrences $\{\widehat{k}_n = 1\}$. This proportion is shown in Figure 1 for the sample sizes $n = 50$ and $n = 150$: it increases clearly with the sample size as well as with θ (*corr* in the axis), whereas it seems independent from the innovations variance σ^2. Proportions very close to 1 are observed for $n = 150$ and θ larger than 0.5.

To evaluate the finite sample behaviour of hPE we calculate its MISE and plot in Figure 2 the quantity $ln(1000MISE)$ in order to improve the visualization; also, the vertical axis scale goes from 0 to 12 to make easier a visual comparison with the case $AR(1)$. Figure 2 (top) shows that the estimation error of hPE is small in all the treated cases and it descreases with the sample size, while it increases with θ and σ^2. Indeed, the maximum value of the MISE of hPE in our simulation study is very small and this is not surprising since a $MA(1)$ process is a favourable case for hPE.

4.2. $AR(1)$ *process*

When observations are generated by a $AR(1)$ process the real order of development is infinity. To look at the behaviour of $\widehat{k}_n$ we consider the proportion (that is the relative frequency) of the occurrences $\{\widehat{k}_n > k_n/2\}$. These proportions

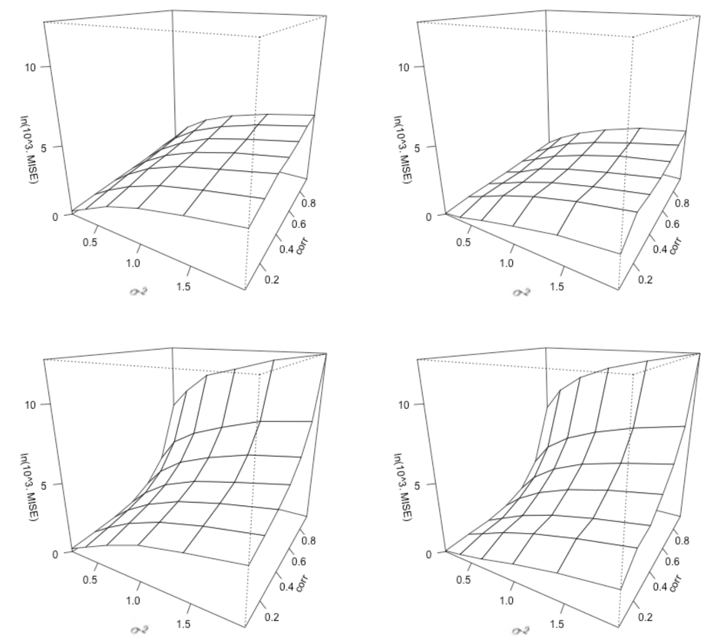

Fig 2. *$ln(1000MISE)$ of hPE in the $MA(1)$ case (top) and $AR(1)$ case (down), with $n = 50$ (left) and $n = 150$ (right)*

are very small, they usually decrease when the innovations variance σ^2 and n increase. This is expectable since, in the case of a $AR(1)$ process, sequence of Fourier coefficients is decreasing, and the choice of γ_n implies that the truncation index will generally be small in this specific case. The unique exception is observed if ϕ takes values close to 1. In this case, the decrease of the Fourier coefficients sequence is so slow that the truncation index assumes larger values. We recall that, with the choice of a large c, $\widehat{k}_n$ will reach the biggest integer j such that the square of the associated estimated Fourier coefficient $\widehat{\varphi}^2_{j,n}$ is larger than the mean of the first $(k_n + 1)$ squared estimated Fourier coefficients.

As shown in Figure 2, the MISE of hPE is clearly higher in the $AR(1)$ case than in the $MA(1)$ case. A possible explanation is that the spectral density of a $AR(1)$ does not admit a finite development with respect to the projection basis (so this is a very unfavorable case for hPE) leading to a less accurate estimation by hPE. Moreover, as expected, the MISE increases with σ^2 and ϕ (*corr* in the axis) and decreases with the sample size.

4.3. Comparison among estimators

In this section, we compare hPE to the classical projection estimator (PE) and to the kernel estimator by examining their evaluated MISEs. In order to summarize the results, we plot the MISE Percentage Increase (MPI) when using hPE with respect to the compared estimator (CE):

$$MPI := 100 * (\text{MISE(CE)} - \text{MISE(hPE)})/\text{MISE(hPE)}$$

A value of $MPI = 100 * x$ is equivalent to MISE(CE) $= (x + 1)$ MISE(hPE). So MPI is positive when hPE has a better performance than the compared estimator CE.

As for the comparison between hPE and the classical projection estimator, Figure 3 (top) shows MPI, for the usual sample sizes, in the $MA(1)$ case. For all the considered values of the innovations variance σ^2 and of the parameter θ, MPI is positive, meaning that hPE has always smaller MISE than the classical PE.

Two phenomena have to be explained here. Firstly, the MPI increases with n. This is due to the fact that the truncation index $\widehat{k}_n$ of hPE reaches very often, as shown in Figure 1, the right order of the development. Contemporaneously k_n - that also appears in the classical PE - increases with the sample size. Secondly, for n larger than 80, MPI usually increases with θ: infact, the first two estimated Fourier coefficients are estimations of quantities respectively proportional to $1+\theta^2$ and 2θ and the following estimated Fourier coefficients are estimations of 0. So, when θ is larger, the threshold γ_n (the quadratic mean of the estimated Fourier coefficients) is likely to be greater than all the estimated Fourier coefficients but the two first ones.

Let us consider now the $AR(1)$ case. The behaviour of MPI is illustrated in

Figure 3 (down). We recall that the MISE of hPE has the form:

$$\mathbb{E}\|\widehat{\varphi}_n - \varphi\|^2 = \mathbb{E}\left(\sum_{j=0}^{\widehat{k}_n} (\widehat{\varphi}_{j,n} - \varphi_j)^2\right) + \mathbb{E}\left(\sum_{j>\widehat{k}_n} \varphi_j^2\right).$$

If $\widehat{k}_n = k_n$, the difference between the MISEs of the classical PE and of hPE is zero. If $\widehat{k}_n < k_n$, such a difference is given by:

$$\text{MISE(PE)} - \text{MISE(hPE)} = \mathbb{E}\left(\sum_{j=\widehat{k}_n+1}^{k_n} (\widehat{\varphi}_{j,n} - \varphi_j)^2\right) - \mathbb{E}\left(\sum_{j=\widehat{k}_n+1}^{k_n} \varphi_j^2\right). \quad (7)$$

Contrary to the previous MA case, MPI decreases with increasing ϕ (*corr* in Figures). Actually, if ϕ is small the information "lost" by hPE with respect to classical PE, that is $\mathbb{E}\left(\sum_{j=\widehat{k}_n+1}^{k_n} \varphi_j^2\right)$, is not important since φ_j is proportional to ϕ^j, and the MPI is clearly positive. If ϕ is large, the truncation index of hPE $\widehat{k}_n$ remains small and - since the Fourier coefficients φ_j are big even for j larger than $\widehat{k}_n$ - the negative term $\mathbb{E}\left(\sum_{j=\widehat{k}_n+1}^{k_n} \varphi_j^2\right)$ in (7) is larger. This quantity is hardly compensated by the "gain" of variance $\mathbb{E}\left(\sum_{j=\widehat{k}_n+1}^{k_n} (\widehat{\varphi}_{j,n} - \varphi_j)^2\right)$ which does not depend on the value of ϕ.

To compare hPE to a kernel spectral density estimator, we estimate the density through the function *spectrum* in R (see R Development Core Team 2015), that uses the modified Daniell kernel by default reporting a bandwidth taken from Bloomfield (2000, p.191). To generate the weights in the linear smoother (see also Shumway and Stoffer 2008), R allows for a repeated use (by the argument *spans* that we set equal to (7,5,8)) of the Daniell kernel so that the resulting kernel will be approximately normal, see Bloomfield (2000, p.195).
In the comparison between hPE and kernel spectral density estimator, we can observe the same phenomena than those observed for the comparison between hPE and classical PE. The main result remains that the hPE, for all the considered cases, has a smaller MISE than the kernel estimator, as Figure 4 shows.

5. Discussion

We consider the nonparametric functional estimation problem that is related to the methodology of thresholding projection and we propose a generalization of the estimation developped in Bosq and Blanke (2007). This estimator has good asymptotic properties but it depends on unknown decreasing parameters of the mixing coefficient sequence, and - as already said - there is not exist an estimator for $\alpha-mixing$ coefficients so far in literature. Moreover, the threshold embedded in its construction depends on a constant that needs to be chosen by the user and on the sample size. The estimator hPE, that we propose in this

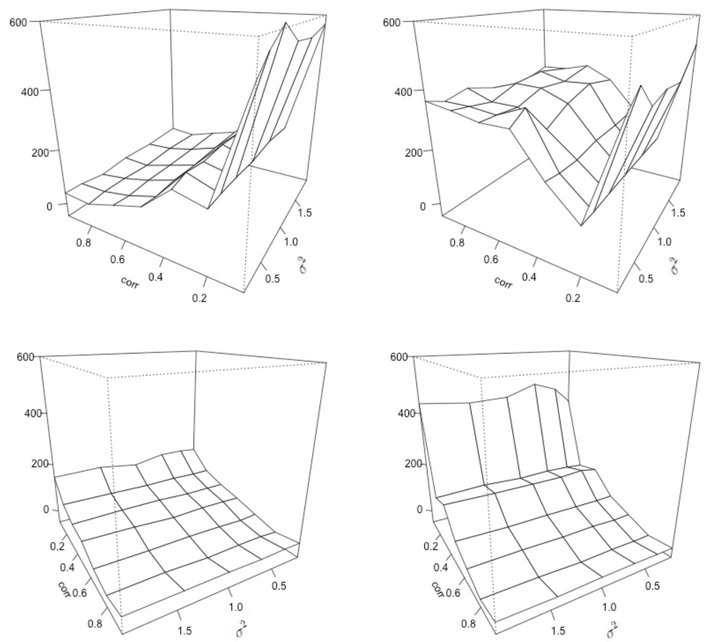

FIG 3. *MISE percentage increase (MPI) of classical PE versus hPE in the $MA(1)$ case (top) and $AR(1)$ case (down), with $n = 50$ (left) and $n = 150$ (right)*

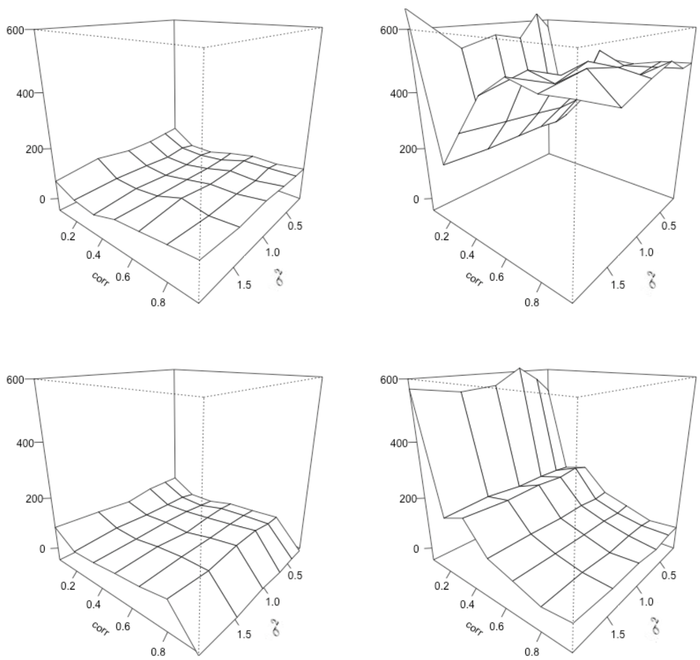

FIG 4. *MISE percentage increase (MPI) of the kernel estimator versus hPE in the $MA(1)$ case (top) and $AR(1)$ case (down), with $n = 50$ (left) and $n = 150$ (right)*

work, addresses these two issues by reducing drastically the dependence from unknown parameters and by increasing the data-driven feature. Moreover, we show that hPE holds good asymptotic properties too, similarly to the estimator in Bosq and Blanke (2007). However, above all, we propose a new data driven choice for the threshold γ_n appearing in hPE definition. Our proposal aims to keep the most part of the available information descarding the estimated Fourier coefficients expected to be smaller than the quadratic mean of the first k_n estimated ones. So, just like for the other compared estimators, the user only has to choose a unique sequence of numbers (the - maximum - dimension of the projection space in the case of the - thresholding - projection estimator, or the bandwidth in the case of the kernel estimator), and nothing else. When the function to estimate is the spectral density, the proposed estimator is shown - by simulations - to reduce often drastically the MISE in comparison with that of the classical projection estimator and the kernel estimator.

Our estimator takes advantage of the fact that it is composed by the estimated Fourier coefficients with the highest absolute values. To this goal, we consider the quadratic mean of the estimated Fouries coefficients as a summary of their sample distribution. Then we keep in the construction of the estimator all the estimated Fourier coefficients larger, in absolute value, than this quadratic mean. Of course, to summarize a distribution we could use another location parameter, such as the mean, the trimmed mean, the median and so on, that could be used instead of the quadratic mean (as shown in Blanke et al., 2012). This could lead to many associated estimators analogous to the one we present in this work.

Finally, this new data driven choice for the threshold γ_n could allow to construct a goodness-of-fit test following Munk *et al.* (2011) that extends the test proposed for dependent α-mixing data with k fixed by Ignaccolo (2004).

Appendix: proofs

5.1. Proof of Proposition 3.1

A result from Bosq (1996, Theorem 1.3, p.25) will be used in this proof and the following ones.

Lemma 5.1. *(***Bosq's inequality)Let** Y_t **an** $\alpha-$**mixing centred real-valued process such that** $\sup_{1\le t\le n}\|Y_t\|_\infty \le b$**. Then** $\forall q \in \mathbb{N}\cap\left[1;\frac{n}{2}\right], \quad \forall\varepsilon>0$**,**

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}Y_i\right|>\varepsilon\right)\le 4\exp\left(\frac{-\varepsilon^2 q}{8b^2}\right)+22q\sqrt{1+\frac{4\|Y_t\|_\infty}{\varepsilon}}\alpha\left(\left\lfloor\frac{n}{2q}\right\rfloor\right) \qquad (8)$$

where $\lfloor x\rfloor$ **is the integer part of** x**.**

Now, for the case **(A)**, let us consider $c>0$ and show that for n large enough

$\sqrt{\frac{\sum_{j=0}^{k_n} \widehat{\varphi_{j,n}}^2}{k_n+1}} > c\sqrt{\frac{\log(n)}{n^\beta}}$ almost surely. We have

$$P\left(\sqrt{\frac{\sum_{j=0}^{k_n} \widehat{\varphi_{j,n}}^2}{k_n+1}} > c\sqrt{\frac{\log(n)}{n^\beta}}\right) \geq P\left(|\widehat{\varphi_{j_0,n}}| > c\sqrt{(k_n+1)\log(n)/n^\beta}\right)$$
$$\geq 1 - P\left(|\widehat{\varphi_{j_0,n}}| \leq c\sqrt{(k_n+1)\log(n)/n^\beta}\right)$$

But $k_n = o\left(\frac{n^\beta}{\log(n)}\right)$ then $(k_n+1)\log(n)/n^\beta \to 0$ and for n large enough, $c\sqrt{\frac{(k_n+1)\log(n)}{n^\beta}} < \frac{|\varphi_{j_0}|}{2} \neq 0$. It follows that $P\left(\sqrt{\frac{\sum_{j=0}^{k_n} \widehat{\varphi_{j,n}}^2}{k_n+1}} > c\sqrt{\frac{\log(n)}{n^\beta}}\right) \geq 1 - P\left(|\widehat{\varphi_{j_0,n}} - \varphi_{j_0}| \geq \frac{|\varphi_{j_0}|}{2}\right)$.

From Bosq's inequality, for $q = 1, 2, \ldots, \lfloor n/2 \rfloor$, $P\left(\sqrt{\frac{\sum_{j=0}^{k_n} \widehat{\varphi_{j,n}}^2}{k_n+1}} > c\sqrt{\frac{\log(n)}{n^\beta}}\right) \geq 1 - (A_q + B_q)$ with $A_q = 4\exp\left(\frac{-\varphi_{j_0}^2 q}{64M^2}\right)$ and $B_q = 22q\sqrt{1 + \frac{16M}{|\varphi_{j_0}|}}\alpha\left(\left\lfloor \frac{n-k_n}{2q} \right\rfloor\right)$. With $q = \lfloor n^\beta \rfloor$, it exists $\lambda > 0$, $\lambda' > 0$ such that $A_q \underset{+\infty}{\sim} 4\exp\left(-\lambda n^\beta\right)$ and $B_q \underset{+\infty}{\sim} \lambda' n^{-\delta(1-\beta)+\beta}$. $\delta > \frac{1+\beta}{1-\beta}$ entails that $\sqrt{\frac{\sum_{j=0}^{k_n} \widehat{\varphi_{j,n}}^2}{k_n+1}} > c\sqrt{\frac{\log(n)}{n^\beta}}$ is true almost surely for n large enough with the lemma of Borel-Cantelli.

In the case **(G)** we have

$$P\left(\sqrt{\frac{\sum_{j=0}^{k_n} \widehat{\varphi_{j,n}}^2}{k_n+1}} > c\sqrt{\frac{\log^{2\Gamma}(n)}{n}}\right) \geq 1 - P\left(|\widehat{\varphi_{j_0,n}}| \leq c\sqrt{(k_n+1)\log^{2\Gamma}(n)/n}\right)$$

For n large enough, $P\left(\sqrt{\frac{\sum_{j=0}^{k_n} \widehat{\varphi_{j,n}}^2}{k_n+1}} > c\sqrt{\frac{\log^{2\Gamma}(n)}{n}}\right) \geq 1 - P\left(|\widehat{\varphi_{j_0,n}} - \varphi_{j_0}| \geq \frac{|\varphi_{j_0}|}{2}\right)$. Analogously to the case (A), from Bosq's inequality, for $q = 1, 2, \ldots, \lfloor n/2 \rfloor$, we have

$$P\left(\sqrt{\frac{\sum_{j=0}^{k_n} \widehat{\varphi_{j,n}}^2}{k_n+1}} > c\sqrt{\frac{\log^{2\Gamma}(n)}{n}}\right) \geq 1 - (A'_q + B'_q)$$

with $A'_q = 4\exp\left(\frac{-\varphi_{j_0}^2 q}{64M^2}\right)$ and $B'_q = 22q\sqrt{1 + \frac{16M}{|\varphi_{j_0}|}}\alpha\left(\left\lfloor \frac{n-k_n}{2q} \right\rfloor\right)$. With $q = \left\lfloor \frac{n}{\log^{2\Gamma-1}(n)} \right\rfloor$, for n large enough, $\forall \lambda > 0, \lambda' > 0, \varepsilon \in]0;1[$, $A'_q \leq 4\exp\left(-\lambda n^{1-\varepsilon}\right)$ and $B'_q \leq \lambda' \frac{n^{1-\log^{2(\Gamma-1)}(n)b/(2+\varepsilon)}}{\log^{2\Gamma-1}(n)}$. This entails that $\sqrt{\frac{\sum_{j=0}^{k_n} \widehat{\varphi_{j,n}}^2}{k_n+1}} > c\sqrt{\frac{\log^{2\Gamma}(n)}{n}}$ almost surely for n large enough with the lemma of Borel-Cantelli.

5.2. Proof of Proposition 3.2

Let introduce the event A_n as

$$A_n = \bigcup_{j=0}^{k_n} \{|\widehat{\varphi}_{j,n}| \geq \gamma_n\}.$$

An interesting property of A_n is described in the following Lemma (for its proof see Bosq and Blanke 2007, p. 82).

Lemma 5.2. *Let $\varphi \in H$ such that $\exists j_0 \in \mathbb{N} : \varphi_{j_0} \neq 0$.*
Then, for n large enough,

$$A_n \text{ holds } a.s.$$

In the following, we consider the event $C_n = \left\{\gamma_n = c\sqrt{\frac{\log(n)}{n^\beta}}\right\}$ in the case **(A)** and $C_n = \left\{\gamma_n = c\sqrt{\frac{\log^{2\Gamma}(n)}{n}}\right\}$ in the case **(G)**, as well as the event $B_n = \{A_n \bigcap C_n\}$. Since, for n large enough, Proposition 3.1 assures that C_n holds almost surely, we have that B_n holds almost surely too.

Now, we start to analyze the behaviour of $\widehat{k}_n$. If $\exists\ K$ such that $\varphi \in \mathcal{G}_0(K)$,

$$\begin{aligned}
\mathbb{P}(\widehat{k}_n \neq K) &= \mathbb{P}(\widehat{k}_n < K) + \mathbb{P}(\widehat{k}_n > K) \\
&= \mathbb{P}\left(\left\{\widehat{k}_n < K\right\} \cap B_n\right) + \mathbb{P}\left(\left\{\widehat{k}_n < K\right\} \cap B_n^c\right) \\
&\quad + \mathbb{P}\left(\left\{\widehat{k}_n > K\right\} \cap B_n\right) + \mathbb{P}\left(\left\{\widehat{k}_n > K\right\} \cap B_n^c\right).
\end{aligned}$$

$\mathbb{P}(B_n^c)$ is larger than the second term and the fourth one since $\left\{\widehat{k}_n < K\right\} \cap B_n^c$ and
$\left\{\widehat{k}_n > K\right\} \cap B_n^c$ are included in B_n^c. Then

$$\mathbb{P}(\widehat{k}_n \neq K) = \mathbb{P}\left(\left\{\widehat{k}_n < K\right\} \cap B_n\right) + \mathbb{P}\left(\left\{\widehat{k}_n > K\right\} \cap B_n\right) + 2\mathbb{P}\left(B_n^c\right)$$

with $2\mathbb{P}(B_n^c)$ which is the general term of a convergent series since Lemma 5.2 was obtained using Borel-Cantelli's lemma. For n large enough, $k_n > K$ so

$$\left\{\widehat{k}_n > K\right\} \cap B_n \Rightarrow \bigcup_{j=K+1}^{k_n} \{|\widehat{\varphi}_{j,n}| > \gamma_n\}.$$

Since φ is an element of $\mathcal{G}_0(K)$, it follows that $\forall j > K, \varphi_j = 0$. So

$$\bigcup_{j=K+1}^{k_n} \{|\widehat{\varphi}_{j,n}| > \gamma_n\} = \bigcup_{j=K+1}^{k_n} \{|\widehat{\varphi}_{j,n} - \varphi_j| > \gamma_n\}$$

$$= \bigcup_{j=K+1}^{k_n} \left\{ \left| \frac{1}{n-\nu(j)} \sum_{i=1}^{n-\nu(j)} h_j(X_i,\ldots,X_{i+\nu(j)}) - \mathbb{E}\left(h_j(X_i,\ldots,X_{i+\nu(j)})\right) \right| > \gamma_n \right\}.$$

Now, introducing $Y_{i,j} = h_j(X_i,\ldots,X_{i+\nu(j)}) - \mathbb{E}\left(h_j(X_i,\ldots,X_{i+\nu(j)})\right)$ and recalling that $\nu(j) \leq j \leq k_n$ by assumption, Bosq's inequality with $q \in \mathbb{N} \bigcap \left[1; \frac{n-\nu(j)}{2}\right]$ entails that

$$\begin{aligned} \mathbb{P}\left(\left\{\widehat{k}_n > K\right\} \cap B_n\right) &= \mathbb{P}\left(\bigcup_{j=K+1}^{k_n} \left\{ \left| \frac{1}{n-\nu(j)} \sum_{i=1}^{n-\nu(j)} Y_{i,j} \right| > \gamma_n \right\}\right) \\ &\leq (k_n - (K+1)) \left[4\exp\left(-\frac{\gamma_n^2 q}{32M^2}\right) + 22q\sqrt{1+\frac{8M}{\gamma_n}}\,\alpha\left(\left\lfloor \frac{n-k_n}{2q} \right\rfloor\right)\right] \\ &\leq (k_n - (K+1))(A_{1,n} + A_{2,n}) \end{aligned} \tag{9}$$

where $A_{1,n} = 4\exp\left(-\frac{\gamma_n^2 q}{32M^2}\right)$ and $A_{2,n} = 22q\sqrt{1+\frac{8M}{\gamma_n}}\,\alpha\left(\left\lfloor \frac{n-k_n}{2q} \right\rfloor\right)$.

Note that $\alpha\left(\left\lfloor \frac{n-\nu(j)}{2q} \right\rfloor\right)$ that should appear in $A_{2,n}$ has been majorated by $\alpha\left(\left\lfloor \frac{n-k_n}{2q} \right\rfloor\right)$. This majoration is true because for n large enough $k_n \geq \nu(j)$ (since $k_n \to \infty$) and α is a decreasing function.

Hereafter we distinguish the two cases related to the α-mixing process.

Case (A): arithmetically mixing

Let $q = \lfloor n^\beta \log(n) \rfloor$ where $0 < \beta < 1$ (so that $q \in \mathbb{N} \bigcap \left[1; \frac{n-\nu(j)}{2}\right]$ for n large enough).

Making explicit α, q and γ_n in the inequality (9), one has $\mathbb{P}\left(\left\{\widehat{k}_n > K\right\} \cap B_n\right) = \mathcal{O}\left(n^{1-c^2\log(n)/32M^2}\right) + \mathcal{O}\left(n^{1+5\beta/4-\delta(1-\beta)}\log(n)^{1/2-\delta}\right)$. Since by assumption $\delta > \frac{2+5\beta/4}{1-\beta}$, then $\exists \zeta_1 > 1 :\ \mathbb{P}\left(\left\{\widehat{k}_n > K\right\} \cap B_n\right) = \mathcal{O}\left(\frac{1}{n^{\zeta_1}}\right)$.

Finally Borel-Cantelli's lemma entails that $\widehat{k}_n \leq K$ almost surely.

Now we want to show that $\widehat{k}_n \geq K$ almost surely. For this goal, let us consider the event $\left\{\widehat{k}_n < K\right\}$.

Since $\gamma_n \to 0$ and $|\varphi_K| \neq 0$, we deduce that, for n large enough, $\frac{|\varphi_K|}{2} > \gamma_n$ and

$$\left\{\widehat{k}_n < K\right\} \cap B_n \Rightarrow \left\{|\widehat{\varphi}_{K,n}| < \frac{|\varphi_K|}{2}\right\}$$

so that

$$\mathbb{P}\left(\left\{\widehat{k}_n < K\right\} \cap B_n\right) \leq \mathbb{P}\left(|\widehat{\varphi}_{K,n} - \varphi_K| > \frac{|\varphi_K|}{2}\right).$$

The previous inequality is equivalent to

$$\mathbb{P}\left(\left\{\widehat{k}_n < K\right\} \cap B_n\right) \leq \mathbb{P}\left(\left|\frac{1}{n-\nu(K)}\sum_{i=1}^{n-\nu(K)} Y_{i,K}\right| > \frac{|\varphi_K|}{2}\right)$$

where $Y_{i,K} = h_K(X_i,\ldots,X_{i+\nu(K)}) - \mathbb{E}\left(h_K(X_i,\ldots,X_{i+\nu(K)})\right)$. Applying the Bosq's inequality with $q \in \mathbb{N} \cap \left[1; \frac{n-\nu(K)}{2}\right]$ one has

$$\mathbb{P}\left(\left\{\widehat{k}_n < K\right\} \cap B_n\right) \leq 4\exp\left(-\frac{\varphi_K^2 q}{128M^2}\right) + 22q\sqrt{1+\frac{16M}{|\varphi_K|}}\alpha\left(\left\lfloor\frac{n-\nu(K)}{2q}\right\rfloor\right).$$

Using again $q = \lfloor n^\beta \log(n)\rfloor$ and $\gamma_n = \sqrt{\frac{\log(n)}{n^\beta}}$ (as precised in Section 2), one has $\mathbb{P}\left(\left\{\widehat{k}_n < K\right\} \cap B_n\right) = \mathcal{O}\left(\exp\left(-\frac{\varphi_K^2 n^\beta}{128M^2}\log(n)\right)\right) + \mathcal{O}\left(n^{\beta-\delta(1-\beta)}\log(n)^{1-\delta}\right)$ so that $\exists\, \zeta_2 > 1$ such that $\mathbb{P}\left(\left\{\widehat{k}_n < K\right\} \cap B_n\right) = o\left(\frac{1}{n^{\zeta_2}}\right)$ since $\delta > \frac{2+5\beta/4}{1-\beta} > \frac{1+\beta}{1-\beta}$.

In conclusion, $\exists\, \zeta > 1$ such that $\mathbb{P}\left(\left\{\widehat{k}_n \neq K\right\}\right) = o\left(\frac{1}{n^\zeta}\right)$ (e.g. $\zeta = min(\zeta_1, \zeta_2)$). This entails that $\widehat{k}_n = K$ a.s. for n large enough.

Case (G): geometrically mixing

In this case, let $q = \left\lfloor\frac{n}{\log^\Gamma n}\right\rfloor$, $\Gamma > 1$ (so that $q \in \mathbb{N} \bigcap \left[1; \frac{n-\nu(j)}{2}\right]$ for n large enough).

Let us consider the terms in the inequality (9): we have $(k_n - (K+1))A_{1,n} = \mathcal{O}\left(n\exp\left(-\frac{c^2}{32M^2}\log^\Gamma n\right)\right) = \mathcal{O}\left(n^{1-c^2\log(n)^{\Gamma-1}/32M^2}\right)$ and for $0 < \varepsilon < 1$, $(k_n - (K+1))A_{2,n} = \mathcal{O}\left(n^{9/4-b/(2+\varepsilon)\log^{\Gamma-1}(n)}\right)$. Hence $\exists\, \zeta > 1$ such that $(k_n - (K+1))A_{2,n} = o\left(\frac{1}{n^\zeta}\right)$.

Borel-Cantelli's lemma implies that $\widehat{k}_n \leq K$ almost surely.

Now we want to show that $\widehat{k}_n \geq K$ almost surely. Using the same ideas as previously, we have that $\exists\, \zeta > 1$ such that $\mathbb{P}\left(\left\{\widehat{k}_n < K\right\} \cap B_n\right) = o\left(\frac{1}{n^\zeta}\right)$.

We conclude that $\widehat{k}_n \geq K$ almost surely.

These results entail that $\widehat{k}_n = K$ almost surely in the (G) case too.

5.3. Proof of Theorem 3.3

Case (A): arithmetically mixing

We want to get a majoration for the quantity $\mathbb{E}\|\widehat{\varphi}_n - \varphi\|^2$ when the right order of development of φ with respect to the projection basis is K. First, we remark that $\mathbb{I}_{\widehat{k}_n = K} + \mathbb{I}_{\widehat{k}_n \neq K} = 1$, so that

$$\mathbb{E}\|\widehat{\varphi}_n - \varphi\|^2 = \mathbb{E}(\|\widehat{\varphi}_n - \varphi\|^2 \mathbb{I}_{\widehat{k}_n = K}) + \mathbb{E}(\|\widehat{\varphi}_n - \varphi\|^2 \mathbb{I}_{\widehat{k}_n \neq K}). \qquad (10)$$

Denoting $\widehat{\varphi}_{n,K} := \sum_{j=0}^{K} \widehat{\varphi}_{j,n} e_j$ we have

$$\mathbb{E}(\|\widehat{\varphi}_n - \varphi\|^2 \mathbb{1}_{\widehat{k}_n = K}) = \mathbb{E}\|\widehat{\varphi}_{n,K} - \varphi\|^2 = \sum_{j=0}^{K} (\widehat{\varphi}_{j,n} - \varphi_j)^2.$$

Now, to get a majoration of $\mathbb{E}(\|\widehat{\varphi}_n - \varphi\|^2 \mathbb{1}_{\widehat{k}_n \neq K})$, we observe that $\|\widehat{\varphi}_n - \varphi\|^2 \leq 2\|\widehat{\varphi}_n\|^2 + 2\|\varphi\|^2$. But $\|\widehat{\varphi}_n\|^2 \leq (k_n+1)M^2$ and $\|\varphi\|^2 \leq (K+1)M^2$. So $\|\widehat{\varphi}_n - \varphi\|^2 \leq 2(k_n + K + 2)M^2$ and

$$\mathbb{E}(\|\widehat{\varphi}_n - \varphi\|^2 \mathbb{1}_{\widehat{k}_n \neq K}) \leq \mathbb{P}(\widehat{k}_n \neq K)\mathcal{O}(k_n).$$

Therefore the Equation (10) becomes

$$\mathbb{E}\|\widehat{\varphi}_n - \varphi\|^2 \leq \mathbb{E}\left(\sum_{j=0}^{K} (\widehat{\varphi}_{j,n} - \varphi_j)^2\right) + \mathbb{P}(\widehat{k}_n \neq K)\mathcal{O}(k_n).$$

Proposition 3.2 entails that, for n large enough,

$$\mathbb{P}(\widehat{k}_n \neq K) \leq \mathcal{O}\left(n^{1-c^2\log(n)/32M^2}\right) + \mathcal{O}\left(n^{1+5\beta/4-\delta(1-\beta)}\log(n)^{1/2-\delta}\right)$$

so $\mathbb{P}(\widehat{k}_n \neq K)\mathcal{O}(k_n) = o\left(\frac{1}{n}\right)$ since we assumed $\delta > \frac{3+\frac{5\beta}{4}}{1-\beta}$, $k_n = o(n)$.

To obtain an upper bound for $\sum_{j=0}^{K} (\widehat{\varphi}_{j,n} - \varphi_j)^2$, let define $Y_{j,i} := h_j(X_i, \ldots, X_{i+\mu(j)}) - \varphi_j$. Then we have

$$\mathbb{E}\left(\sum_{j=0}^{K} (\widehat{\varphi}_{j,n} - \varphi_j)^2\right) = \sum_{j=0}^{K} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{l=1}^{n} |cov(Y_{j,i}, Y_{j,l})|.$$

Using the inequality in Rio (2000, p.9), we deduce that, for $i \neq l$,

$$|cov(Y_{j,i}, Y_{j,l})| \leq 2 \int_0^{|l-i|^{-\delta}} Q_{j,i}(u) Q_{j,l}(u) du$$

where the functions $Q_{j,i}$ and $Q_{j,l}$ stand respectively for the quantile functions associated to the variables $Y_{j,i}$ and $Y_{j,l}$. So

$$|cov(Y_{j,i}, Y_{j,l})| \leq 8M^2 |l-i|^{-\delta}$$

and denoting $\mathbb{C}\,(j,j) := \frac{1}{n^2} \sum_{i=1}^{n} \sum_{l=1}^{n} |cov(Y_{j,i}, Y_{j,l})|$ we have

$$\mathbb{C}\,(j,j) \leq \frac{1}{n^2} \sum_{(i,l) \in 1\ldots n,\ i \neq l} 8M^2 |l-i|^{-\delta} + \frac{1}{n^2} \sum_{i=1}^{n} var(Y_{j,i}).$$

Then, with $r = |i-l| \neq 0$,

$$\mathbb{C}\,(j,j) \quad \leq \quad \frac{16M^2}{n}\sum_{r=1}^{n-1} r^{-\delta} + \frac{1}{n^2}\sum_{i=1}^{n} var(Y_{j,i})$$

and, using the inequality $\sum_{r=1}^{n-1} r^{-\delta} \leq 1 + \int_1^{n-1} x^{-\delta}dx$, we get

$$\mathbb{C}\,(j,j) = \mathcal{O}\left(\frac{1}{n^{1-\delta}}\right) + \frac{1}{n^2}\sum_{i=1}^{n} var(Y_{j,i}).$$

Moreover, $var(Y_{j,i}) < 4M^2$, so $\frac{1}{n^2}\sum_{i=1}^{n} var(Y_{j,i}) = \mathcal{O}\left(\frac{1}{n}\right)$. Since $0 < \beta < 1$ and $\delta > \frac{2+\frac{5\beta}{4}}{1-\beta}$ we have $\delta > 2$ and we conclude that

$$\mathbb{E}\|\varphi_{n,K} - \varphi\|^2 = \sum_{j=0}^{K} \mathbb{C}\,(j,j) = \mathcal{O}\left(\frac{1}{n}\right)$$

that is $n\mathbb{E}\|\widehat{\varphi}_n - \varphi\|^2 = O(1)$, and the proof is completed.

Case (G): geometrically mixing
The first difference with respect to the arithmetical case is in the majoration of $\mathbb{P}(\widehat{k}_n \neq K)$. We deduce from Proposition 3.2 that $\mathbb{P}(\widehat{k}_n \neq K) \leq o\left(\frac{1}{n^2}\right)$. This entails that $\mathbb{E}\|\widehat{\varphi}_n - \varphi\|^2 = o\left(\frac{1}{n}\right)$.

Another difference is observed when majorationg $\mathbb{C}(j,j)$. In fact, from Rio (2000, p.9) we have

$$|cov(Y_{j,i}, Y_{j,l})| \leq 2\int_0^{a\exp(-b|l-i|)} Q_{j,i}(u)Q_{j,l}(u)du \leq 8M^2 a\exp(-b|l-i|)$$

with the same previous notations. Hence

$$\mathbb{C}\,(j,j) \quad \leq \quad \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n} 8M^2 a\exp(-b|l-i|) + \frac{8M^2a}{n} \leq \frac{8M^2a}{n}\left(1 + \frac{2}{1-\exp(-b)}\right)$$

and this inequality allows us to complete the proof, analogously to the case (A).

5.4. Proof of Proposition 3.4

If $\varphi \in \mathcal{G}_1$, respectively under the assumption $\delta > \frac{2+5\beta/4}{1-\beta}$ for the arithmetical case, for n large enough, then

$$\widehat{k}_n > j \quad a.s.$$

for all j such that $\varphi_j \neq 0$. This comes from the fact that such a j is smaller than the true order of development and we showed in Proposition 3.2 that $\widehat{k}_n$

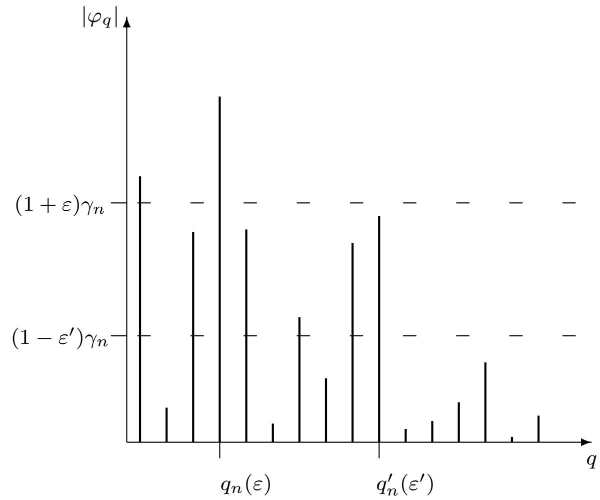

FIG 5. *An example of* $q_n(\varepsilon)$ *and* $q'_n(\varepsilon')$

is, for n large enough, almost surely larger than the true order of development. This entails

$$\liminf_{n\to\infty} \widehat{k}_n > j \quad a.s.;$$

since $\varphi \in \mathcal{G}_1$ there is an infinity of such j, so we conclude that, for n large enough, $\widehat{k}_n \to \infty$.

5.5. *Proof of Theorem 3.5*

First, we introduce a new parameter depending on the functional parameter φ:

$$q(\eta) := min\left\{q \in N : |\varphi_j| \le \eta, \quad \forall j > q\right\}, \quad \eta > 0$$

and denote $q_n(\varepsilon)$ and $q'_n(\varepsilon')$ the integers $q((1+\varepsilon)\gamma_n)$ and $q((1-\varepsilon')\gamma_n)$ for all $\varepsilon > 0$ and $\varepsilon' \in]0,1[$; an example is illustrated in Figure 5.
Note that $q(\eta)$ is defined through the real values of the Fourier coefficients φ_j, and it is well defined because the sequence $(\varphi_j)_{j\in\mathbb{N}}$ converges to 0 when $n \to \infty$ since $\sum_{i=0}^{\infty} \varphi_j^2 < \infty$ $(\varphi \in H)$. Nevertheless, intuitively we can expect that $\widehat{k}_n$ will take a value between $q_n(\varepsilon)$ and $q'_n(\varepsilon')$: indeed the following property confirms this intuition (almost surely for n large enough).

Lemma 5.3. *For every* $\varepsilon > 0$, $\varepsilon' \in]0;1[$, *if* $k_n > q_n(\varepsilon)$ *and if, in the case* **(A)**, $\delta > \frac{2+5\beta/4}{1-\beta}$ *then for* n *large enough, for both cases:*

$$q_n(\varepsilon) \le \widehat{k}_n \ \le \ \min\left(q'_n(\varepsilon'), k_n\right) \qquad a.s.$$

Proof of Lemma 5.3 If $q_n(\varepsilon) = 0$, the first inequality is obvious. If $q_n(\varepsilon) > 0$, by construction of $q_n(\varepsilon)$ and since $k_n > q_n(\varepsilon)$ we have

$$\begin{aligned} \mathbb{P}\left(\left\{\widehat{k}_n < q_n(\varepsilon)\right\} \cap B_n\right) &\leq \mathbb{P}\left(\left|\widehat{\varphi}_{q_n(\varepsilon),n} - \varphi_{q_n(\varepsilon)}\right| > \varepsilon\gamma_n\right) \\ &\leq \mathbb{P}\left(\left|\frac{1}{n-\nu(q_n(\varepsilon))}\sum_{i=1}^{n-\nu(q_n(\varepsilon))} Y_{i,q_n(\varepsilon)}\right| > \varepsilon\gamma_n\right) \end{aligned}$$

with the previous notations.

Now, we use the Bosq's inequality for both the cases (A) and (G).

Case (A): arithmetically mixing
With $q = \lfloor n^\beta \log(n) \rfloor$,

$$\mathbb{P}\left(\left\{\widehat{k}_n < q_n(\varepsilon)\right\} \cap B_n\right) \leq \mathcal{O}\left(n^{\frac{c^2\varepsilon^2\log(n)}{32M^2}}\right) + \mathcal{O}\left(n^{1+5\beta/4-\delta(1-\beta)}\right).$$

But $\delta > \frac{2+5\beta/4}{1-\beta}$ so we deduce that

$$\exists\, \zeta > 1: \ \mathbb{P}\left(\left\{\widehat{k}_n < q_n(\varepsilon)\right\} \cap B_n\right) = o\left(n^{-\zeta}\right).$$

For n large enough, B_n is verified almost surely and the inequality

$$\mathbb{P}\left(\widehat{k}_n < q_n(\varepsilon)\right) \leq \mathbb{P}\left(\left\{\widehat{k}_n < q_n(\varepsilon)\right\} \cap B_n\right) + \mathbb{P}\left(B_n^c\right)$$

entails that the event $\left\{\widehat{k}_n \geq q_n(\varepsilon)\right\}$ is true almost surely for n large enough by the lemma of Borel-Cantelli.

For the other inequality $\widehat{k}_n \leq \min(q_n'(\varepsilon'), k_n)$, we use once more the Bosq's inequality with $q = \lfloor n^\beta \log(n) \rfloor$ and obtain that

$$\mathbb{P}\left(\left\{\widehat{k}_n > \min(q_n'(\varepsilon'), k_n)\right\} \cap B_n\right) = \mathcal{O}\left(n^{1-\left(\frac{c^2\varepsilon'^2\log(n)}{32M^2}\right)} + n^{1+5\beta/4-\delta(1-\beta)}\right).$$

We conclude again with the lemma of Borel-Cantelli.

Case (G): geometrically mixing
Let $q := \left\lfloor \frac{n}{\log^\Gamma n} \right\rfloor$ in the Bosq's inequality. For n large enough and the previous notations,

$$\mathbb{P}\left(\left\{\widehat{k}_n < q_n(\varepsilon)\right\} \cap B_n\right) = \mathcal{O}\left(n^{\frac{-c^2}{32M^2}\log^{\Gamma-1}(n)}\right) + \mathcal{O}\left(n^{\frac{5}{4}-2b\log^{\Gamma-1}(n)}\right).$$

So

$$\exists\, \zeta > 1: \ \mathbb{P}\left(\left\{\widehat{k}_n < q_n(\varepsilon)\right\} \cap B_n\right) = \mathcal{O}\left(n^{-\zeta}\right)$$

and $\widehat{k}_n \geq q_n(\varepsilon)$ almost surely for n large enough.

The last inequality $\widehat{k}_n \leq \min(q'_n(\varepsilon'), k_n)$ is shown analogously to the case (A), taking $q = \left\lfloor \frac{n}{\log^{\Gamma} n} \right\rfloor$ in the Bosq's inequality. Hence the proof of Lemma 5.3 is completed.

The convergence rates for $MISE(\widehat{\varphi}_n)$ when $\varphi \in \mathcal{G}_1$ are based on the inequalities provided by the following lemma.

Lemma 5.4. *Let denote* $m_\varphi^2 := \inf_{j\geq 0} var(h_j(X_1, \dots, X_{1+\nu(j)}))$ *and suppose* $m_\varphi^2 \neq 0$.
If k_n *is large enough, then in the case* **(A)***, with* $\delta > 1$

$$m_\varphi^2 \frac{(q_n(\varepsilon)+1)}{n} + \sum_{j>\min(q'_n(\varepsilon'),k_n)} \varphi_j^2 \ \leq \ \mathbb{E}\|\widehat{\varphi}_n - \varphi\|^2$$

and

$$\mathbb{E}\|\widehat{\varphi}_n - \varphi\|^2 \ \leq \ \frac{5M^2}{\delta - 1} \frac{(\min(q'_n(\varepsilon'), k_n) + 1)}{n} + \sum_{j>q_n(\varepsilon)} \varphi_j^2 + o\left(\frac{1}{n}\right);$$

in the case **(G)***,*

$$m_\varphi^2 \frac{(q_n(\varepsilon)+1)}{n} + \sum_{j>\min(q'_n(\varepsilon'),k_n)} \varphi_j^2 \ \leq \ \mathbb{E}\|\widehat{\varphi}_n - \varphi\|^2$$

and

$$\mathbb{E}\|\widehat{\varphi}_n - \varphi\|^2 \ \leq \ \frac{5M^2 a}{1-\exp(-b)} \frac{(\min(q'_n(\varepsilon'), k_n) + 1)}{n} + \sum_{j>q_n(\varepsilon)} \varphi_j^2 + o\left(\frac{1}{n}\right).$$

Proof of Lemma 5.4 ***Case (A): arithmetically mixing***
First, let $A_n := \left\{q_n(\varepsilon) \leq \widehat{k}_n \leq \min(q'_n(\varepsilon'), k_n)\right\}$. Then, the mean integrated square error is equal to

$$\mathbb{E}\|\widehat{\varphi}_n - \varphi\|^2 = \mathbb{E}\left(\left(\sum_{j=0}^{\widehat{k}_n} (\widehat{\varphi}_{j,n} - \varphi_j)^2 + \sum_{j>\widehat{k}_n} \varphi_j^2\right)(\mathbb{1}_{A_n} + \mathbb{1}_{\overline{A_n}})\right)$$

Note that

$$\mathbb{E}\left(\left(\sum_{j>\widehat{k}_n} \varphi_j^2\right)\left(\mathbb{1}_{A_n} + \mathbb{1}_{\overline{A_n}}\right)\right) \leq \sum_{j>q_n(\varepsilon)} \varphi_j^2 + \|\varphi\|^2 \mathbb{P}(\overline{A_n}).$$

But $\mathbb{P}(\overline{A_n}) = o\left(\frac{1}{n}\right)$ because the assumptions of Lemma 5.3 are fulfilled ($\delta > \frac{2+5\beta/4}{1-\beta}$ for the case (A)), so

$$\mathbb{E}\left(\left(\sum_{j>\widehat{k}_n} \varphi_j^2\right)\left(\mathbb{1}_{A_n} + \mathbb{1}_{\overline{A_n}}\right)\right) \leq \sum_{j>q_n(\varepsilon)} \varphi_j^2 + o\left(\frac{1}{n}\right).$$

Analogously,

$$\mathbb{E}\left(\sum_{j=0}^{\widehat{k}_n}(\widehat{\varphi}_{j,n}-\varphi_j)^2\mathbb{1}_{A_n}\right)\leq\sum_{j=0}^{\min(q_n'(\varepsilon'),k_n)}\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}Y_{j,i}\right)^2.$$

Thanks to the inequality in Rio (2000, p. 9) with $\delta>\frac{2+5\beta/4}{1-\beta}>1$,

$$\mathbb{E}\left(\sum_{j=0}^{\widehat{k}_n}(\widehat{\varphi}_{j,n}-\varphi_j)^2\mathbb{1}_{A_n}\right)\leq\frac{4M^2}{n}\frac{\min(q_n'(\varepsilon'),k_n)+1}{\delta-1}.$$

Moreover,

$$\mathbb{E}\left(\sum_{j=0}^{\widehat{k}_n}(\widehat{\varphi}_{j,n}-\varphi_j)^2\mathbb{1}_{\overline{A_n}}\right)\leq 4(k_n+1)M^2\mathbb{P}(\overline{A_n}).$$

We deduce that

$$\exists\zeta>1:\ \mathbb{E}\left(\sum_{j=0}^{\widehat{k}_n}(\widehat{\varphi}_{j,n}-\varphi_j)^2\mathbb{1}_{\overline{A_n}}\right)=o\left(\frac{k_n}{n^\zeta}\right)$$

and, for n large enough,

$$\mathbb{E}\left(\sum_{j=0}^{\widehat{k}_n}(\widehat{\varphi}_{j,n}-\varphi_j)^2\mathbb{1}_{\overline{A_n}}\right)<\frac{M^2}{\delta-1}\frac{k_n+1}{n}.$$

This last step completes the part of the proof dedicated to the upper bound.

For the lower bound, we have

$$\mathbb{E}\left(\sum_{j=0}^{\widehat{k}_n}(\widehat{\varphi}_{j,n}-\varphi_j)^2\right)\geq\mathbb{E}\left(\sum_{j=0}^{\widehat{k}_n}(\widehat{\varphi}_{j,n}-\varphi_j)^2\mathbb{1}_{A_n}\right)\geq\frac{(q_n(\varepsilon)+1)m_\varphi^2}{n}$$

and

$$\mathbb{E}\left(\sum_{j>\widehat{k}_n}\varphi_j^2\right)\geq\mathbb{E}\left(\sum_{j>\widehat{k}_n}\varphi_j^2\mathbb{1}_{A_n}\right)\geq\sum_{j>\min(q_n'(\varepsilon'),k_n)}\varphi_j^2.$$

Case (G): geometrically mixing

The only difference with respect to the previous case arises from the following inequality

$$\mathbb{E}\left(\sum_{j=0}^{\widehat{k}_n}(\widehat{\varphi}_{j,n}-\varphi_j)^2\mathbb{1}_{A_n}\right)\leq\frac{4M^2}{n}\frac{a\min(q_n'(\varepsilon'),k_n)+1}{1-\exp(-b)}$$

where $\frac{a}{1-\exp(-b)}$ appears instead of $\frac{1}{\delta-1}$ (for the case (A)).
Hence the proof of Lemma 5.4 is completed.

From the previous inequalities in Lemma 5.4 it is possible to derive the convergence rate of $MISE(\widehat{\varphi}_n)$ and conclude the proof of Theorem 3.5.
In both cases (A) and (G), the idea is to split the biggest term of the previous inequalities, that is $\sum_{j>q_n(\varepsilon)} \varphi_j^2$, in two parts. More precisely, we give an upper bound for $\sum_{j=q_n(\varepsilon)+1}^{k_n} \varphi_j^2$: in the case **(A)**,

$$\sum_{j=q_n(\varepsilon)+1}^{k_n} \varphi_j^2 \leq c^2 k_n (1+\varepsilon)^2 \frac{\log(n)}{n^\beta},$$

while in the case **(G)**,

$$\sum_{j=q_n(\varepsilon)+1}^{k_n} \varphi_j^2 \leq c^2 k_n (1+\varepsilon)^2 \frac{\log^{2\Gamma}(n)}{n}.$$

References

[1] Antoniadis, A. (2007). Wavelet methods in statistics: Some recent developments and their applications. *Statistics Surveys*, 1, 16-55.

[2] Aubin, J.B. (2005). Estimation localement suroptimale de la densité par projection tronquée. *Ann. I.S.U.P.*, 49 (2-3), 3-17.

[3] Aubin, J.B. (2006). Comportement asymptotique d'estimateurs de la densité par projection tronqués. *C. R. Math. Acad. Sci. Paris*, 342 (11), 873–876.

[4] Aubin, J.B. & Leoni-Aubin, S. (2008). Projection Density Estimation under a m-sample Semiparametric Model. *Computational Statistics and Data Analysis*, 52 (5), 2451–2468.

[5] Blanke, D., Gabriel E. & Josselin D. (2012) Comparing new adaptive and robust estimators of location. *Ann. I.S.U.P.*, 56(2-3), 6585.

[6] Bleuez, J. & Bosq, D. (1976). Conditions nécessaires et suffisantes de convergence de l'estimateur de la densité par la méthode des fonctions orthogonales. *C. R. Acad. Sci. Paris*, 282 (17, Aiii), A1023–A1026.

[7] Bloomfield, P. (2000). *Fourier analysis of time series*, 2nd ed., Wiley Series in Probability and Statistics, New York.

[8] Bosq, D. (2002). Estimation localement suroptimale et adaptative de la densité. *C. R. Math. Acad. Sci. Paris*, 334 (7), 591–595.

[9] Bosq, D. (2005). Estimation suroptimale de la densité par projection. *Canadian Journal of Statistics*, 33 (1), 21–37.

[10] Bosq, D. & Blanke, D. (2007). *Inference and prediction in large dimensions.* Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester.

[11] Bosq D. & Lecoutre, J.P. (1987). *Théorie de l'Estimation Fonctionnelle.* Economica, Paris.

[12] Carbon, M. (1984). Sur l'estimation asymptotique d'une classe de parametres fonctionnels. *Publ. Inst. Statist. Univ. Paris*, 29 (1), 13–26.

[13] Cencov, N. N. (1962). Evaluation of an unknown distribution density from observations. *Soviet math.*, 3, 1559–1562.

[14] Delecroix, M. & Protopopescu, C. (2001). Regression series estimators: the MISE approach. *J. Nonparametr. Statist.*, 13 (4), 453–483.

[15] Devroye, L. (1987). *A course in density estimation.* Progress in Probability and Statistics, vol. 14, Birkhauser Boston Inc., Boston, MA.

[16] Devroye, L. & Gyorfi, L. (1985). *Nonparametric density estimation, The L1 view.* Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics, John Wiley and Sons Inc., New York.

[17] Diggle, P. J. & Hall, P. (1986). The selection of terms in an orthogonal series density estimator. *J. Amer. Statist. Assoc.*, 81 (393(, 230–233.

[18] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. & Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.*, 24 (2), 508–539.

[19] Efromovich, S. (1996). Adaptive orthogonal series density estimation for small samples. *Computational Statistics and Data Analysis*, 22 (6), 599–617.

[20] Efromovich, S. (1999). *Nonparametric curve estimation.* Springer Series in Statistics, Springer-Verlag, New York.

[21] Efromovich, S. (2010). Orthogonal series density estimation. *WIREs Computational Statistics*, 2 (4), 467–476.

[22] Efron, B. & Tibshirani, R. (1996). Using specially designed exponential families for density estimation. *Ann. Statist.*, 24 (6), 2431–2461.

[23] Eubank, R. L. & Speckman, P. (1990). Curve fitting by polynomial-trigonometric regression. *Biometrika*, 77 (1), 1–9.

[24] Hart, J. D. (1985). On the choice of a truncation point in Fourier series density estimation. *J. Statist. Comput. Simulation*, 21 (2), 95–116.

[25] Hart, J. D. (1997). *Nonparametric smoothing and lack-of-fit tests.* Springer Series in Statistics, Springer-Verlag, New York.

[26] Ignaccolo, R. (2004). Goodness-of-fit tests for dependent data, *Nonparametric Statistics*, 16 (1–2), 19–38.

[27] Kolmogorov, A. N. & Fomin, S. V. (1957). *Elements of the theory of functions and functional analysis.* Vol. 1. Metric and normed spaces, Graylock Press, Rochester, N. Y., Translated from the first Russian edition by Boron, L.F.

[28] Kronmal, R. & Tarter, M. (1968). The estimation of probability densities and cumulatives by Fourier series methods. *J. Amer. Statist. Assoc.* **63**, 925-952.

[29] Krylov, A. N. (1955). *Collected works, XII, part one.* Various works, Izdat. Akad. Nauk SSSR, Moscow.

[30] Lepski, O. V., Mammen, E. & Spokoiny, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors, *Ann. Statist.* ,25 (3), 929–947.

[31] McDonald, D.J., Cosma Rohilla Shalizi, Schervish, M. (2015). Estimating beta-mixing coefficients via histograms. *arXiv:1109.5998v2*

[32] Munk, A., Stockis, J.-P., Valeinis, J., & Giese, G. (2011). Neyman smooth goodness-of-fit tests for the marginal distribution of dependent data, *Annals of the Institute of Statistical Mathematics*, 63 (5), 939-959.

[33] Nadaraya, E.A. (1989). *Nonparametric estimation of probability densities and regression curves, Mathematics and its applications.* Soviet series, Kluwer cademic Publisher.

[34] Picard, D. & Tribouley, K. (2000). Adaptive confidence interval for pointwise curve estimation, *Ann. Statist.*, 28 (1), 298–335.

[35] Prakasa Rao, B. L. S. (1983). *Nonparametric functional estimation.* Probability and Mathematical Statistics, Academic Press Inc, New York.

[36] R Development Core Team (2015). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

[37] Rio, E. (2000). *Théorie Asymptotique des Processus Aléatoires Faiblement Stationnaires.* Springer.

[38] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA*, 42, 43-47.

[39] Sansone, G. (1959). *Orthogonal functions.* Revised English ed. Translated from the Italian by Diamond, A. Pure and Applied Mathematics, Vol. IX, Interscience Publishers, Inc., New York.

[40] Shumway R.H. & Stoffer, D.S. (2008). *Time series analysis and its applications. With R examples*, 2nd ed., Springer Texts in Statistics, Springer, New York.

[41] Silverman, B.W. (1986). *Density estimation for statistics and data analysis.* Monographs on Statistics and Applied Probability, Chapman and Hall, London.

[42] Souare, M. (2008). Estimation suroptimale de la densité spectrale par projection adaptative: application a l'estimation de l'ordre d'une moyenne mobile, *C. R. Math. Acad. Sci. Paris*, 346 (17-18), 999–1002.

[43] Tarter, M. & Lock, M.D. (1993). *Model-free curve estimation.* Monographs on Statistics and Applied Probability, vol. 56, Chapman and Hall, New York.

[44] Thompson, J.R. & Tapia, R.A. (1990). *Nonparametric function estimation, modeling, and simulation.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.